Title
Children Are Not Just Noisy Adults: Disentangling Noise and Bias in Numerical Estimation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Park, Hyekyung
Opfer, John

Publication Date
2023

Peer reviewed

Children Are Not Just Noisy Adults: Disentangling Noise and Bias in Numerical Estimation

Hyekyung Park (park.2766@osu.edu)
Department of Psychology, The Ohio State University
1835 Neil Avenue, Columbus, OH 43210 USA

John Opfer (opfer.7@osu.edu)
Department of Psychology, The Ohio State University
1835 Neil Avenue, Columbus, OH 43210 USA

Abstract

There are ongoing debates whether logarithmic compression in number-to-space mapping reflects logarithmic encoding of large numbers (bias) or uncertainty about numeric value (noise). We tested these two hypotheses by disentangling the effect of bias and noise. When 80 adults and 80 4- to 7-year-olds were asked to estimate the number of dots on a number line, both children and adults were more logarithmic on 0-100 than 0-30 problems. Internal noise explained some of the variance in logarithmicity, but only for children. We then examined the wisdom of crowds effect by comparing accuracy of children's mean estimate with accuracy of each adult's estimates. As predicted, children as a crowd were not as accurate as individual adults, indicating that noise is not the only source of children's errors. Generally, increasing the size of a crowd also had a smaller effect on 0-100 than 0-30 problems, indicating that inaccuracy on 0-30 problems is likely due to noise. The present study provides evidence that bias and noise have an additive effect on logarithmic compression and that children's logarithmicity reflects bias in number representations, not just noise.

Keywords: wisdom of crowds effect; number-line estimation; cognitive development

Introduction

Inaccurate judgments entail errors that consist of noise, bias, or both (Kahneman, Sibony, & Sunstein, 2021). As noise and bias have additive effects on error, error rates across age groups actually reveal little about cognitive development. Does accuracy improve with age because children become less noisy, less biased, or both? To answer this question, we focused on developmental changes in estimation. A novel feature of our study was to revisit Galton's (1907) *wisdom of crowds effect*, wherein the mean (and median) estimated weight of an ox was as accurate as the best individual estimate. Like Galton, we wondered if the average estimate of a crowd of children, where noise is cancelled out, is more accurate than that of any individual adult? As we will see, prominent theories of numerical representation have different implications for the wisdom of crowds effect.

In our study, we asked children and adults, not to estimate the weight of an ox, but to estimate the numerosity of dots on a number-line. Number-line estimation is preferable to Galton's method because humans and other species spontaneously map numbers to space (Adachi, 2014; de Hevia, Izard, Coubart, Spelke, & Streri, 2014; de Hevia & Spelke, 2010; Drucker & Brannon, 2014; Lourenco, 2010; Rugani, Vallortigara, Priftis, & Regolin, 2015), even if they cannot report a specific number (e.g., because they are not fluent with the names of large numbers). Number-to-space mapping has been examined extensively using the number-line estimation task, where participants are asked to estimate the position of a numerosity or numerical value on a horizontal line flanked by 0 and a larger number (Siegler & Opfer, 2003; Dehaene, Izard, Spelke, & Pica, 2008). Number-line studies typically show that differences between small numbers tend to be overestimated and that differences between larger numbers are underestimated, much like numbers on a logarithmic ruler. Some researchers propose that this logarithmic compression reflects an evolved bias in encoding numerical information (Dehaene et al., 2008; Siegler & Opfer, 2003; Yuan, Prather, Mix, & Smith, 2020). For a hungry animal, the differences between 1 and 10 pieces of food matter more than the differences between 101 and 110 pieces of food. In contrast, others suggest that noise and uncertainty result in logarithmic compression, perhaps because the bounded number line results in the truncation of responses (Rips, 2013). In this paper, we aim to test these two hypotheses about bias versus noise.

Logarithmic Encoding of Numerical Information

Previous studies have shown that young children show logarithmic compression in number-line estimates. With education and numerical experience, children's estimates become more linear. This *log-to-linear shift* occurs first for small and familiar numbers, resulting in the co-existence of logarithmic and linear representations at any given age (Siegler, Thompson, & Opfer, 2009). For example, Kim and Opfer (2017) showed that kindergarteners estimated 0-30 number lines linearly, but 0-100 number lines logarithmically. Similarly, second graders estimated 0-100 number lines linearly, but 0-1000 number lines logarithmically (Siegler & Opfer, 2003). Third graders estimated 0-1000 number lines linearly, but 0-10,000 number lines logarithmically (Thompson & Opfer, 2010). Thus, at any given age, children are likely to show a mixture of logarithmic and linear responses.

To quantify the degree of logarithmicity in number line estimates, Anobile, Cicchini, and Burr (2012) proposed a mixed log-linear model (MLLM):

$$y = a\left(\lambda \frac{U}{\ln(U)} \ln(x) + (1-\lambda)x\right),$$

where y denotes the estimate of a target number x on a number line with an upper-bound U. a is a scaling parameter. λ is

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

the index of logarithmic compression in estimates, measuring the relative contribution of log representations in comparison with that of linear representations. When estimates are perfectly logarithmic, λ equals 1. When estimates are perfectly linear, λ equals 0. Studies have shown that λ decreases with age and increases with number range of number line (Lee, Kim, Opfer, Pitt, & Myung, 2022). Lee et al. (2022) examined logarithmic compression in numerosity estimates on 0-50, 0-100, 0-200, and 0-400 number line. Kindergarteners to second graders were most logarithmic regardless of the range of number line, followed by third to seventh graders and adults. In addition, both children and adults were more logarithmic on 0-400 than 0-50 number lines.

In addition to the effect of numeric range, estimates change from trial to trial depending on the information available from previous trials. Kim and Opfer (2018) examined the source of logarithmic compression by analyzing logarithmicity of estimates in trial-by-trial basis. In their study, logarithmicity of each trial was calculated across participants. For example, logarithmicity of first trial was calculated by collapsing first estimates of all participants. When 5- to 6-year-old children and adults were asked to estimate the number of dots on the 0-30 number line, both children ($\lambda = 1$) and adults ($\lambda = 0.70$) were logarithmic on the first trial. Nevertheless, estimates of adults became linear over trials ($\lambda = 0.18$ for the last trial) while the last trial of children ($\lambda = 1$) was as logarithmic as their first trial. Dynamics in the number-line estimation further supports that the natural psychological scaling of numbers is logarithmic, but context of the number-line task such as numeric range or sequential effect help to suppress the intuitive logarithmic representation for more accurate representation.

Effect of Uncertainty on Number-to-Space Mapping

Against the logarithmic encoding hypothesis, an alternative approach proposes that numbers are encoded linearly, but the noise of representation increases with numerical value. As a result, the uncertainty that covaries with numerical value drives logarithmic compression in number line estimation task (Cantlon, Cordes, Libertus, & Brannon, 2009). According to this linear-scalar hypothesis, increasing the uncertainty of stimuli will increase logarithmicity of estimates regardless of the numeric range of the number line. Chesney and Matthews (2013) manipulated uncertainty by presenting numbers in either decimal form (e.g., 2272 or 4960) or exponential form (e.g., $.018 \times 10^{4.5}$ or $.025 \times 10^{4.5}$). Adults were more logarithmic when numbers were presented in unfamiliar form (exponential) than in familiar form (decimal).

More recently, Cicchini, Anobile, Chelli, Arrighi, and Burr (2022) examined whether internal noise of representation and external noise of stimuli drive compression of estimates. To examine the effect of internal noise on logarithmic compression of estimates, internal noise was quantified using Dispersion Index (DI) which is based on response variance. Like Kim and Opfer (2018), Cicchini et al. (2022) calculated logarithmic component λ and DI on a trial-by-trial basis. Both λ value and DI was found to decrease over trials, resulting in a positive correlation between logarithmicity and DI. The researchers then examined the effect of external noise on compression of estimates by presenting color line estimation task. In the task, color patches were presented instead of dot arrays. External noise was added either at the high end or at the low end of the color line. The result showed that estimates were logarithmic when external noise was added at the upper end, whereas estimates were exponential when the noise was added at the lower end. From these findings, Cicchini et al. (2022) proposed that internal and external noise drive compression of estimates, rather than logarithmic encoding of number.

Wisdom of Crowds

An effective tool for decomposing error into noise and bias is the wisdom of crowds effect (Surowiecki, 2005). When estimates are unbiased, estimation errors of individuals cancel when averaged, leaving an accurate mean estimate (Galton, 1907; Lorenz, Rauhut, Schweitzer, & Helbing, 2011). In his study, Galton (1907) analyzed the guesses in weight-judging competition where competitors had to estimate the weight of the ox. The result showed that the mean of estimates across competitors was as accurate as the best guesser, which differed 1lb from the actual weight of the ox (Galton, 1907; Hooker, 1907). After Galton (1907)'s study, numerous studies have shown that the average of individuals' estimates tends to be more accurate than the estimate of individuals or even experts (Hommes, Sonnemans, Tuinstra, & Van de Velden, 2005; Lorenz et al., 2011; Lorge, Fox, Davitz, & Brenner, 1958; Yaniv & Milyavsky, 2007).

We simulated the wisdom of crowds effect and effect of crowd size on accuracy of estimates given different assumptions about the internal scaling of number. Our simulation shows that logarithmic and linear-scalar encoding can be distinguished based on the wisdom of crowds effect (Figure 1). Given typical logarithmic compression for large numbers ($\lambda = 0.5$) and constant variability (SD = 10), the wisdom of crowds effect would not be observed (Figure 1C left). In addition, increasing the crowd size from 10 to 50 would not improve accuracy of mean estimates (Figure 1A). In contrast, if noise covaries with numerical value (SD = $.4 \times$ target number) with no compression ($\lambda = 0.0$), both a wisdom of crowds effect (Figure 1C right) and an effect of crowd size would be observable (Figure 1B). Thus, the wisdom of crowds effect provides a valid and novel method for testing competing theories of number representation.

The Current Study

In this paper, we propose that the effects of numeric value and uncertainty on logarithmic compression are not mutually exclusive. Indeed, education and numeric experience decrease both bias and uncertainty about numbers, leading adults' estimates to be less biased and less noisy than children's. Like Cicchini et al. (2022), we manipulated external noise of stimuli and measured internal noise using DI. Additionally, we ex-

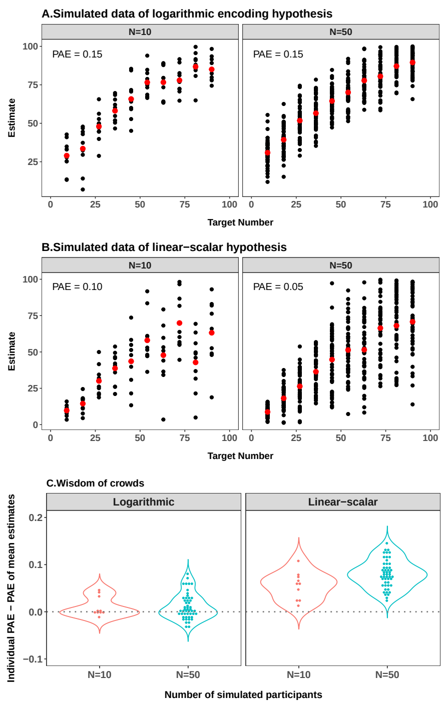

Figure 1: A. Simulated data assuming logarithmic representation with constant response variance. B. Simulated data assuming linear representation with scalar variability. A, B. Left panels show the estimates of 10 simulated participants. Right panels show the estimates of 50 simulated participants. Red dots indicate mean estimates of each target number. Percent absolute error (PAE) was calculated based on mean estimates. Increasing the number of simulated participants decreased PAE of mean estimates only for linear-scalar model. C. Difference between each individual's PAE and PAE of the crowd's mean estimate. Wisdom of crowds effect was most salient in linear-scalar model, where the average estimate of a large crowd was as accurate as the most accurate individual.

amined the wisdom of crowd effect for small numbers (where bias was expected to be low) and large numbers (where bias was expected to be high).

To examine the effect of external noise, we orthogonally manipulated numerical value and perceptual entropy, an information-theoretic measure of uncertainty. Entropy is defined as a weighted average of the number of bits of information that are required to predict value of the stimulus (DeWind, Bonner, & Brannon, 2020; Shannon & Weaver, 1949; Young & Wasserman, 2001). Perceptual entropy increases with a set size because more items need to be encoded to estimate numerosity. In addition to set size, perceptual entropy is higher when a set consists of more diverse objects compared to homogenous set. For example, an 8-dot array with 8 distinct colors has higher entropy than an 8-dot array with same color (Qu, DeWind, & Brannon, 2022). In the present study, we manipulated perceptual entropy by including two levels of numeric range and two levels of perceptual variability. We included 0-30 and 0-100 number lines which were presented either in single-colored dot arrays or multiple-color dot arrays. By measuring external noise based on perceptual entropy, we aimed to decompose the effect of external noise that covary with numerical value from the effect of numeric value itself.

In addition to perceptual entropy, we investigated the effect of internal noise on logarithmic compression of estimates. We quantified internal noise based on dispersion index (DI) proposed by Cicchini et al. (2022). Cicchini et al. (2022) calculated DI across participants for each trial order to examine the relation between logarithmicity and internal noise. However, calculating DI across participants measures disagreement between participants rather than internal noise of number representation within participant. To overcome this limitation, we calculated DI on a subject by subject basis. Subsequent regression analyses then allowed us to examine the source of logarithmic compression by comparing the relative effects of numerical value, perceptual entropy, and internal noise on estimates.

The second purpose of the present study was to use the wisdom of crowd effect to examine whether children are more erroneous than adults because children are simply noisier than adults or more biased. To investigate this issue, we compared the accuracy of the mean estimate of a small crowd of children, the mean estimate of a large crowd of children, and the accuracy of individual adult estimates. If children's errors were solely based on noise (as we predict for small numbers), the mean estimate of a large crowd of children would be as accurate as the most accurate adults, when accuracy is measured based on percent absolute error (PAE). If children's errors were due to bias (as we predict for large numbers), estimates of individual adults would be more accurate than the mean estimate of a large crowd of children.

Method

Participants

Ninety-six undergraduate students and eighty-four 4- to 7-year-old children participated in the study. The experiment was conducted online for adults. Adults received course credit for their participation. Two participants who took longer than 30 minutes and two participants whose estimates were completely logarithmic (λ = 1) across all conditions were excluded. Twelve participants who replied stimuli did not fit the screen were excluded. A total of eighty adults were included in the analysis (M = 20.81 years, SD = 3.41 years; 47 females, 30 males, 3 nonbinaries). Age of one participant was not available.

Children completed the in-person experiment at schools. Children received a sticker for their participation. Four children who did not complete the experiment were excluded. A total of eighty children were included in the analysis (M = 6.17 years, SD = 0.68 years; 42 females, 38 males).

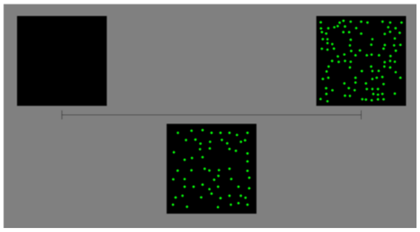

Figure 2: Illustration of a number-line estimation trial

Materials and Procedures

The experiment was conducted using jsPsych library version 6.2.0 (de Leeuw, 2015) for both adults and children. Adults completed the experiment online. Children completed the experiment on a 13-inch laptop. Adults and children completed a nonsymbolic number-line estimation task individually at a computer. The size and the presentation time of each experimental stimulus were equal between adults and children. For adults, we asked whether experimental stimuli fit on the screen. In the task, a 1000px horizontal line was presented on the neutral gray background. The line was flanked by 0 dots on the left end and a fixed number of dots on the right end. Dot arrays were presented on the 300px × 300px black square. Once participants clicked a 30px × 30px white box at the bottom of the screen, a target dot array was presented for 750ms below the number line. Participants were asked to estimate the number of dots by clicking a position on a line (Figure 2).

The entropy of each condition was manipulated based on two levels of numeric range and perceptual variability. Numerical range of number line was either 0 to 30 (small) or 0 to 100 (large). Dot arrays with low perceptual variability were generated using a single color (single). Dot arrays with high perceptual variability were generated by randomly choosing a color for each dot (multi). Dot arrays were generated using MATLAB with Psychophysics Toolbox-3. Color of dots was chosen from HSV color space. Single-colored dot arrays were presented with green dots [HSV: 0.33 1 1]. For multiple-colored dot arrays, hue was randomly chosen for each dot. Saturation and value were fixed to 1.

Participants completed four conditions: single-small, single-large, multi-small, and multi-large. In each condition, 10 target numerosities were chosen to sample non-subitizable numbers evenly. For 0 to 30 number line, target numerosities were 5, 8, 11, 13, 16, 18, 21, 23, 26, and 28. For 0 to 100 number line, target numerosities were 17, 27, 37, 43, 53, 60, 70, 77, 87, and 93. Each target numerosity was presented twice. On half of the trials, the dot size of target stimuli was equal to the dot size of the right-end dots. On the other half of the trials, total surface area covered by dots was equal to the accumulated area of the right-end dots.

Each participant completed 80 trials in total, with 20 trials in each condition. The order of target stimuli was determined by a balanced Latin square, such that each target numerosity was presented twice on each order of the 20 trials in a block. The order of condition was counterbalanced. Instructions were given before each block, and participants started the task with no practice trials.

Analysis

Perceptual entropy of the upper-end dot arrays was calculated using the entropy function in MATLAB, defined as:

$$Entropy = -\sum_{i=1}^{n} p_i \log_2 p_i$$

where p_i is the proportion of individual pixel value belonging to the i^{th} category of value in an array, and n is the total number of pixel values. We calculated entropy for each channel of RGB based on the pixel value and averaged them. The entropy value was 0.04 for 30 single-colored dots, 0.11 for 30 multi-colored dots, 0.10 for 100 single-colored dots, and 0.36 for 100 multi-colored dots.

Internal noise was calculated based on dispersion index (DI) proposed by Cicchini et al. (2022). Researchers first computed response variance as follows:

$$\hat{\sigma}_i^2 = \sum_{j}^{n} \frac{(y_{i,j} - \overline{y_{i,j}})^2}{n} \Big/ x_i$$

where y denotes the estimate of a target number x and n denotes the number of estimates of a target number x. DI was then calculated as follows:

$$DI = \sqrt{\sum_{i}^{N} \hat{\sigma}_i^2 / N}$$

where N denotes the number of target numbers. Unlike Cicchini et al. (2022), the present study includes multiple ranges of number line. Therefore, we scaled raw responses y and target numerosity x by dividing them by the upper bound U to scale DI across multiple upper bounds. Scaled DI of each participant was computed. To calculate scaled DI, we first computed the variance of scaled raw responses $y_{i,j}/U$ for each target numerosity x_i where i denotes numerosity condition ($x_1, x_2, \ldots, x_{10}$) and j denotes counter running of the numerosity x_i. We then normalized each response variance by dividing it by scaled target numerosity x_i/U. We then averaged these normalized variances across ten numerosity stimuli and took their square root to obtain the dispersion index.

Results

We first examined the effects of numeric range of number line and external noise on internal noise. DI of each participant was regressed against numeric range and perceptual entropy of the upper-end dot array. For adults, DI increased with numeric range, b = 0.23, SE = 0.06, $p < .001$. However, the effect of entropy was not significant, b = -0.09, SE = 0.06, p = 0.16. Similarly for children, DI increased with numeric range, b = 0.15, SE = 0.06, $p < .05$, but the effect of entropy was not significant, b = -0.05, SE = 0.06, p = 0.411.

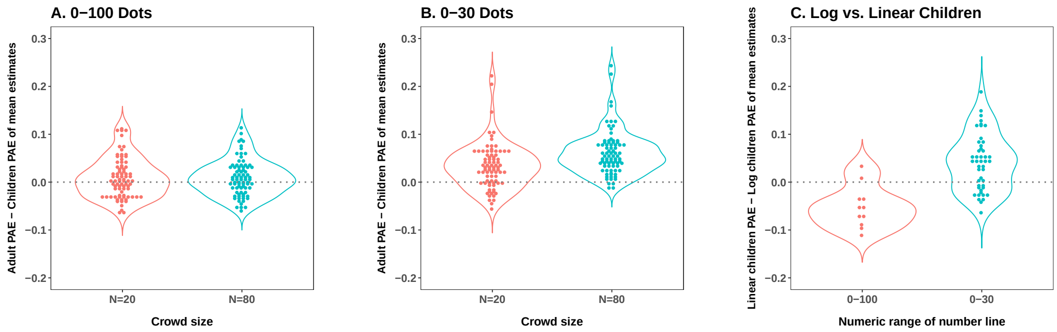

Figure 3: A, B. Difference between adult's individual PAE and PAE of mean estimates of children in single-colored conditions. Negative values indicate adults whose estimates were more accurate than children's mean estimates. Effects of crowd size were larger in small numeric range (0-30) than in larger numeric range (0-100). The difference in the wisdom of crowds effect indicates errors for large numeric range cannot be due to noise alone. C. Difference between linear children's individual PAE and PAE of mean estimates of logarithmic children.

Our main interest was to examine the relative effects of number, internal noise, external noise on logarithmic compression. Across four conditions, children (single-small: $\lambda = 0.34$, single-large: $\lambda = 0.51$, multi-small: $\lambda = 0.24$, multi-large: $\lambda = 0.48$) were more logarithmic than adults (single-small: $\lambda = 0.0$, single-large: $\lambda = 0.19$, multi-small: $\lambda = 0.0$, multi-large: $\lambda = 0.18$). We predicted that if the logarithmic compression is derived by uncertainty, the effects of internal and external noise would be larger than the effect of number on logarithmic component. In contrast, if logarithmic compression in number line estimation is due to bias, the effect of number would be larger. To examine this, logarithmic component (λ) of each participant was regressed against the numeric range of number line, DI, and entropy of the upper-end dot array. For adults, logarithmicity increased with numeric range of number line, b = 0.36, SE = 0.05, $p < .001$. However, we did not find the significant effects of DI, b = 0.01, SE = 0.05, p = 0.912, and entropy, b = -0.04, SE = 0.05, p = 0.416. The results indicate that adults' logarithmic compression of estimates reflects internal bias of number representation, not due to noise. When same regression analysis was conducted for children, children's logarithmicity increased with numeric range of number line, b = 0.21, SE = 0.05, $p < .001$, and DI, b = 22, SE = 0.05, $p < .001$. The effect of entropy was not significant, b = -0.04, SE = 0.05, p = 0.409. Unlike adults, internal noise and numerical range had additive effect on children's compression.

Overall, our results indicated that log-to-linear shift reflects both representational change and reduction in noise. To further investigate the source of children's error, we examined whether children exhibited the wisdom of crowds effect. To compare errors of children with errors of adults, we computed percent absolute error (PAE) as follows:

$$PAE = \left| \frac{Estimate - TargetNumber}{UpperBound} \right|.$$

Higher PAE value indicates less accurate estimation. To calculate errors of children as a crowd, PAE of mean estimates were calculated for each condition in each block. Therefore, we calculated mean estimates of 20 children for each target number and computed PAE of these mean estimates. For adults, PAE was calculated for each participant. We then subtracted children's PAE of mean estimates from adults' individual PAE. For example, children's PAE of mean estimates in single-small condition in the first block was subtracted from each adult's individual PAE of single-small condition done in the first block. We predicted that if children's compression is due to the noisy representation of numbers, then children's PAE of mean estimates would be as low as the PAE of the most accurate adult. If children's compression reflects the logarithmic encoding of numbers, then children's PAE of mean estimates would be higher than the most accurate adult. The result showed that 26.25% of adults in single-small condition, 45% of adults in single-large condition, 36.25% of adults in multi-small condition, and 48.75% of adults in multi-large condition exhibited lower PAE than the PAE of mean estimates of children. Even though children's response noise was canceled out by averaging their estimates, their mean estimates were not as accurate as adults.

Next, we examined if children as a crowd would outperform individual adults when we increase the crowd size. To examine this, PAE of mean estimates of all children were calculated for each condition after collapsing data across blocks. We calculated mean estimates of 80 children for each target number and computed PAE of these mean estimates. When children's PAE of mean estimates was subtracted from adults' individual PAE, 5% of adults in single-small condition, 45% of adults in single-large condition, 18.75% of adults in multi-small condition, and 43.75% of adults in multi-large condition exhibited lower PAE than children's PAE of mean estimates.

Overall, examining the wisdom of crowds effect of children further supported our hypothesis that logarithmic encoding of numbers and uncertainty have additive effects on children's compression. Even after reducing the response variance of children by averaging their estimates, they were more erroneous than the most accurate adult. Interestingly, the effect of the crowd size differed depending on the numeric range of the number line. Increasing crowd size reduced the proportion of adults who were more accurate than children as a crowd on 0-30 number line, suggesting that children's error can be partly explained by noise (Figure 3B). However, the crowd size did not have an effect on 0-100 number line (Figure 3A). This result suggests that larger logarithmicity in larger numeric range reflects the internal bias of number, not due to uncertainty that covaries with numeric value.

We next examined the wisdom of crowds effect among children by comparing children who were completely logarithmic ($\lambda = 1$) with children who were completely linear ($\lambda = 0$). Due to the small N of completely logarithmic children (single-small: 12, single-large: 14, multi-small: 8, multi-large: 11) and linear children (single-small: 19, single-large: 5, multi-small: 27, multi-large: 6) per condition, we combined single-colored conditions with multiple-colored conditions, considering that the effect of perceptual entropy was not significant. When we examined the wisdom of crowds effect among children, 30.43% of children who were completely linear on 0-30 number line had lower PAE than the PAE of mean estimates of children who were completely logarithmic. In 0-100 number line condition, 81.82% of linear children had lower PAE than the PAE of mean estimates of logarithmic children (Figure 3C). Smaller effect of wisdom of crowds in larger numeric range further suggests that logarithmicity that increases with numeric range of number line reflects bias of encoding numerical information and that internal noise alone cannot explain compression in number-to-space mapping.

Discussion

The present study investigated the sources of errors and developmental change by comparing the effects of noise and bias on number-line estimation. Like previous studies, we found that numeric range of number lines (0-30 vs. 0-100) contributed to logarithmic compression of estimates across the life span, and the effect of internal noise decreased with age. These results suggest that children's large number representations are both noisier and more biased than those of adults. Parallel results were found in wisdom of crowds effects, which allow us to disentangle effects of bias versus noise. Against Galton (1907), we found that a crowd of children was not as accurate as any individual adult, further supporting the idea that noise is not the only source of errors. This was particularly true for large numbers. Increasing children's crowd size improved performance on 0-30 but not on 0-100 number lines. The effect of numeric range was also evident when examining the wisdom of crowds effect among children themselves, some of whom provided adult-like (linear) estimates and some of whom provided immature (logarithmic) estimates. On 0-100 number lines, most individual linear children showed lower error rates than the average of a crowd of logarithmic children. On 0-30 number lines, the wisdom of crowds effect was more salient.

Based on these findings, we propose numbers are initially encoded logarithmically. Over development, children learn to minimize both bias and noise for more accurate estimates. If children's errors were solely based on bias, we would not have found the wisdom of crowds effect nor the effect of internal noise on logarithmicity. If noise were the only source of children's logarithmic compression, numeric range of number line would not have affected the logarithmicity of estimates after controlling for the effects of internal and external noise. Further, averaged estimates of children would have been as accurate as the most accurate adult (i.e., the wisdom of crowds). We suggest both noise and bias drive logarithmic compression, based on our intermediate findings that showed the additive effect of internal noise and bias and wisdom of crowds effect which differed depending on the numeric range. A follow up study on symbolic number line task, where Arabic numerals are presented instead of dot arrays, would allow us to generalize the present findings to symbolic number representation.

Examining wisdom of crowds effect allowed us to disentangle noise versus bias by reducing the noise in estimates. However, averaging estimates does not reduce internal noise of number representation. In the present study, we mainly focused on reducing the noise in estimates because previous studies proposed truncation of responses due to the bounded feature of number line could be the source of logarithmic compression (Rips, 2013). Future studies that directly manipulate internal noise would provide further evidence on the additive effects of noise and bias on estimation errors.

The sources of errors in human judgment and decision making has been investigated in a wide variety of contexts (Kahneman et al., 2021; Surowiecki, 2005). In addition to numerical cognition, our findings shed light on cognitive development in general. Estimation or judgment across different situations become more accurate with age. The present study illustrates how the wisdom of crowds effect can be an effective method to investigate developmental changes in human judgment.

References

Adachi, I. (2014). Spontaneous spatial mapping of learned sequence in chimpanzees: evidence for a snarc-like effect. *PLoS One*, *9*(3), e90373.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.

Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on "log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures". *science*, *323*(5910), 38–38.

Chesney, D. L., & Matthews, P. G. (2013). Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks. *Psychonomic Bulletin Review*, *20*(6), 1146–1153.

Cicchini, G. M., Anobile, G., Chelli, E., Arrighi, R., & Burr, D. C. (2022). Uncertainty and prior assumptions, rather than innate logarithmic encoding, explain nonlinear number-to-space mapping. *Psychological Science*, *33*(1), 121–134.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures. *science*, *320*(5880), 1217–1220.

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*, *111*(13), 4809–4813.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, *21*(5), 653–660.

de Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

DeWind, N. K., Bonner, M. F., & Brannon, E. M. (2020). Similarly oriented objects appear more numerous. *Vision*, *20*(4), 4–4.

Drucker, C. B., & Brannon, E. M. (2014). Rhesus monkeys (macaca mulatta) map number onto space. *Cognition*, *132*(1), 57–67.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*(7), 450–451.

Hommes, C., Sonnemans, J., Tuinstra, J., & Van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *The Review of Financial Studies*, *18*(3), 955–980.

Hooker, R. H. (1907). Mean or median. *Nature*, *75*(1951), 487–488.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.

Kim, D., & Opfer, J. E. (2017). A unified framework for bounded and unbounded numerical estimation. *Developmental psychology*, *53*(6), 1088.

Kim, D., & Opfer, J. E. (2018). Dynamics and development in number-to-space mapping. *Cognitive Psychology*, *107*, 44–66.

Lee, S. H., Kim, D., Opfer, J. E., Pitt, M. A., & Myung, J. I. (2022). A number-line task with a bayesian active learning algorithm provides insights into the development of non-symbolic number estimation. *Psychonomic Bulletin Review*, *29*(3), 971–984.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, *108*(22), 9020–9025.

Lorge, I., Fox, D., Davitz, J., & Brenner, M. (1958). A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological bulletin*, *55*(6), 337.

Lourenco, M. R., S. F. abd Longo. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873–881.

Qu, C., DeWind, N. K., & Brannon, E. M. (2022). Increasing entropy reduces perceived numerosity throughout the lifespan. *Cognition*, *225*, 105096.

Rips, L. J. (2013). How many is a zillion? sources of number distortion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1257.

Rugani, R., Vallortigara, G., Priftis, K., & Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science*, *347*(6221), 534–536.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information. *Urbana: University of Illinois Press*, *97*(6), 128–164.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological science*, *14*(3), 237–250.

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education*, *3*(3), 143–150.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Thompson, C. A., & Opfer, J. E. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*, *81*(6), 1768–1786.

Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, *103*(1), 104–120.

Young, M. E., & Wasserman, E. A. (2001). Entropy and variability discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 278.

Yuan, L., Prather, R., Mix, K. S., & Smith, L. B. (2020). Number representations drive number-line estimates. *Child development*, *91*(4), e952–e967.